ROAD DATA UPDATING
USING TOOLS OF MATCHING AND MAP GENERALIZATION

HU Yun gang [a,*], CHEN Jun[b] , LI Zhi lin[c] , ZHAO Ren liang[b]

[a]Beijing University of Civil Engineering and Architecture, 100044, Beijing, China – hyg@bucea.edu.cn
[b]National Geomatics Center of China, Beijing 100044, China
[c]Dept. of Land Surveying and Geo-Informatics The Hong Kong Polytechnic University, 999077, Hong Kong Kowloon, China

Commission IV, WG IV/3

KEY WORDS: Map Generalization, Updating, Road Data, Matching, Selective Omission

ABSTRACT:

It is one of the important ways for GIS data updating based on map generalization. This paper analyzes the main steps for road data updating based on map generalization. As the core of this updating process, matching method considering the levels analyses and selective omission based on mesh density are developed. The approach for road data updating based on these two tools is proposed, which is applied in the project of national GIS data updating at 1:50,000 scale from data at 1:10,000 scale ant the results shows its feasibility.

1. INTRODUCTION

GIS data need to be updated to keep its up-to-date. As the use of GIS is spreading into various fields and our environment changes over time, the demand for updating GIS data is increasing. The old GIS data can be updated using the new data at larger scale based on map generalization. It is one of the important ways for GIS data updating, due to its efficiency, economy and data consistency preserved.

The methods of GIS data updating have been studied by several researchers in recent years, most of which are focused on the update of MRDB (Multiple Representation Database). Kilpeläinen and Sarjakoski（1995）discussed the incremental update based on map generalization in a MRDB. Harrie and Hellström (1999) developed a prototype system for propagating updates from large-scale data to small-scale data. Badard and Lemarie（1999）described a tool of matching for updating, and so on. However, how to update the old GIS data at small scale using the new data at large scale when the MRDB of these data is not built? These approaches mentioned above are still very limited, so new tools of matching and generalization methods are required for this update process. This paper aims at updating GIS data at 1:50,000 using new data at 1:10,000 based on map generalization and focus on studying road data updating.

2. MAIN STEPS FOR UPDATING PROCESS

This updating process can be divided in two main steps. Retrieval of updates from two data at different scales is regarded as the first step. Then the second step is to update the old database with the information received.

The first step consists of two main sub-steps, i.e. data matching and selective omission. Matching is a critical first step for extracting updates and it is to establish the correspondence relationships between geographical objects that represent the same phenomenon in the real world(Gabay and Doytsher，1994；Filin and Doytsher，1999；Walter and Fritsch，1999). The two representations may have very different scales, levels of abstraction and different production time in this study. Selective omission is one of generalization operations for road feature. It is a process to retain more important road features (or to eliminate less important ones), while the essential topological, geometric and semantic characteristics of a road network are preserved(Jiang and Harrie, 2004；Mackaness，1995). The results of matching and selective omission are overlaid and analyzed in order to retrieve updates. The unmatched roads on the data at small scale are regarded as the disappearance of roads. The roads on this database need to be deleted and can be called the disappearance of generalization, which are the counterparts of the roads matched and unselected on the data at large scale. The roads unmatched and selected on the data at large scale are considered as the new or change roads. These updates retrieved are coarse and not taken to update directly the old data, which operations belong to the second step for updating.

The other operations of map generalization for road feature such as simplification, typification and displacement should be consisted in the second step. This step also involves addition and deletion of database operations as well as the preservation spatial integrity. In this step, some operations may be carried out according to the algorithms developed, for example simplification can be applied the algorithm presented by Li-OpenShow (Li and Openshaw, 1992; 1993). However most of the operations are difficult to automatic implement. Especially the operation of displacement need to deal with the specific conflicts to preserve the integrity of the database updated. Nevertheless the handled objects are mostly updates in this step, so these sub-processes can be implemented by interactive operation with the relative few workloads.

* Corresponding author.

Thus the first step is the core of the updating process, and the key is to develop the algorithms for matching and selective omission. Then they are detailed in the next two sections

3. MATCHING BASED ON ANALYSIS OF LEVELS

Beside application in GIS data updating, network matching also plays an important role in image processing, data integration and so on, so considerable efforts have been devoted for network matching. These methods mostly focus on an investigation of algorithm that consists of three types of matching, i.e. node matching, segment matching, and edge matching. The matching measures for nodes matching include distance, the degree of node and the value of spider function (Rosen and Saalfeld, 1985; Saalfeld, 1988). Edge matching can be implemented based on the node and segment matching (Gabay and Doytsher, 1994; Filin and Doytsher, 1999; Lemarie and Raynal, 1996), or directly using of the measures and the buffer tool (Walter and Fritsch 1999, Badard 1998, zhang et al., 2005). However, these approaches mentioned above are focused on matching of data with same or similar scales.

3.1 Analyses for matching levels

Indeed, a road network can be regarded as the composing of line feature and crossroad feature, so the analyses for matching levels are also divided into two parts according to the matching of two type features.

One edge can be divided into segments and several edges can be joined a route. So for line feature, three levels are displayed, and we name segment as the decomposed level object, edge as the basic level objects and route as the abstract level object. As to line feature matching, three matching levels are also formed according to line feature levels (see figure 1).

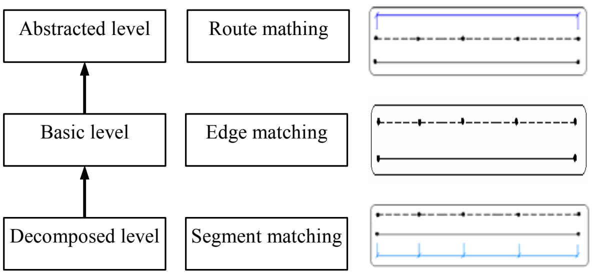


Figure 1. Three matching levels for linear feature

It is difficult for using measures of matching to realize edge matching of many to many. However, for matching of segment or route, the types of matching are either 1:1 or 1:0, due to segments or routes as the results of edges spited or joined. So segment or route matching can be easily confirmed using measures, and edge matching can be obtained based on these two kind matching. Moreover using of segment matching is frequently adopted methods (Xiong and Sperling, 2004). Depending on the criteria of length and direction segments can be split along edges with same scales. Nevertheless, for different scales segments split and their counterparts are not enough equal due to the different abstract levels, and it will affect accuracy of edge matching. It is not feasible using segment matching to obtain edge matching under this condition.

A joined route and its counterpart can be easily built, so route matching is used to represent edge matching in this paper.

Route matching is divided into 1:N and M:N（M≥0, N≥0）matching according to the number of edges on the routes corresponding (see figure 2), where 1(or M) represents number of edges on the route at large scale and N represents number of edges on the other route at small scale. Edge matching can be regarded as the route mapping of one to one, and be induced to route matching.

(a) Route matching of 1:N (b) Route matching of M:N

Figure 2. Two types of route matching

Simple crossroads are all described as one node on two road networks at different scales. However, complex crossroads are represented differently on two road networks at different scales. These crossroads are composed of nodes and edges. So matching of nodes and matching between node and edge are defined. In addition, an end point can be regarded as the decomposed level of a node, and the matching end points will help to build node matching and route matching.

Based on the analysis above, the crossroads corresponding can also be divided into three levels of matching. The end point matching can be regarded as the decomposed matching level, and the node matching as the basic matching level, and the route matching and matching between node and edge as the abstracted matching level, see figure 3.

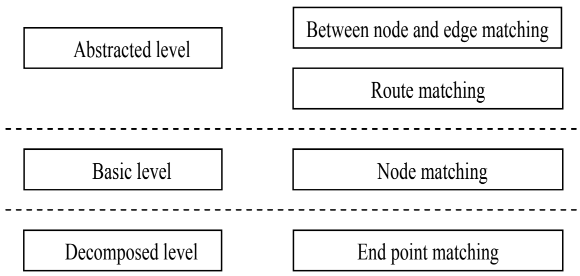


Figure 3. Three matching levels for crossroad feature

3.2 Strategy and methods for matching

According to the principle from the simple to the complex, the method of matching adopts the bottom-up strategy. It is starting with end point matching, then proceeding node matching, and finally ending up route matching and matching between node and edge. This strategy is consistent with the order of matching levels from the decomposed to the abstract. To obtain the higher levels of matching, they need to be transformed into the lower levels of matching, or using the lower levels of matching. For instance, M:N route matching can be transformed 1:N route matching and 1:N route matching can be transformed into 1:1 route matching again by joining edges with different algorithms. In addition, the low level of matching may also need the help from the high level of matching. For example, the part of matching of end point need infer form the node matching. The strategy of matching is shown in figure 4.

For 1:N route matching, the buffers around every edge need to be created firstly. Secondly according to the matching for end points of an edge, the edges corresponding are joined within given buffer, and then the candidate route matching is found. Lastly, the 1:1 route match will be identified by comparing measures including length, Huausdorff distance and semantic attributes of the routes. Accomplished 1:1 route matching, the M:N route matching is processed in succession. The unmatched edges on the data at small scale are to be joined into a route at every node based on the principle of perceptual organization, which describes two edges with the same direction at the intersection. So M:N route matching will transform into the matching of 1:N. The method of 1:N route matching mentioned above can be carried out again.

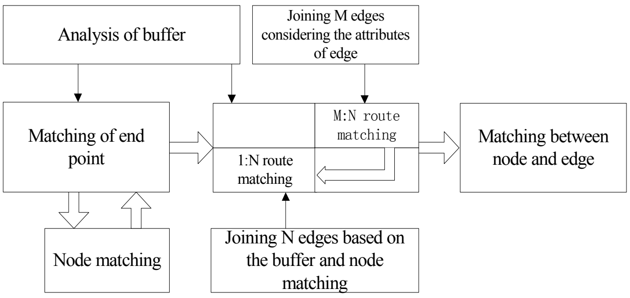


Figure 4. Strategy of matches of objects

Accomplished route matching, the unmatched edges on the data at large scale are filtered. The matching between node and edge can be easily inferred according to its definition mentioned above.

4. SELECTIVE OMISSION BASED ON MESH AND MESH DENSITY

For the selective omission of roads, various techniques already exist. For example, the graph theory principles (Mackaness & Beard, 1993), space syntax (Mackaness, 1995; Jiang & Claramunt, 2004) and self-organizing maps (Jiang & Harrie, 2004) are employed to support generalization of road network. Unfortunately, all these methods failed to consider the distribution density of roads which is an important constraint.
Hu (2007) reveals that selective omission of roads should be retained the density difference besides topological, geometric and semantic properties of the road network. Two distinctive types of density variations are considered, with one across different regions within the same map (e.g. urban and rural areas) and the other across different map scales of the same region. The density conventionally defined as the ratio of total length of roads in a given region to the area of the region. However, this measure is not sufficient for the purpose of map generalization because the local variations of road density over the space are not indicated. So a novel approach of selective omission for roads is used in this study, which is based on mesh and mesh density in a network.

4.1 Concepts of mesh and mesh density

A mesh is defined as a naturally closed region that does not contain any other region in a road network. Naturally, one could consider using area of an arbitrary mesh in the new measure for road density. It is called mesh density. In fact, mesh density is a special case of the road density expressed conventionally and can be described as follows:

$$D = P/A \qquad (1)$$

Where P is the perimeter of a mesh, A is the area of the mesh, and D represents the mesh density. Form this equation we can find that the road stubbles within the mesh boundary are all ignored. The road segments in a network can be classified into two types, i.e. the segments that constitute mesh boundaries and the stubbles lying within the boundaries. Road stubbles can be relatively easily handled according to their geometric and semantic properties. The elimination of any road stubble will not influence the connectivity of network. Moreover, if a mesh needs to be simplified, the road stubbles within the mesh will be likely omitted prior to boundary segments. For these reasons, this simplification is adopted.

The meshes can be classified based on the types of roads on its boundary. A mesh may be bordered by roads of different classes such as main streets and secondary streets. Then meshes are classified based on the bounding road segments. In this method, the class of a mesh is assigned with bounding segment of the lowest ranking. For road generalization based on the mesh density, a mesh can be regarded as the basic unit of selection. With a given set of roads, the mesh density can be mapped out. The mesh density is then used as a constraint to determine which meshes should be treated. Usually, a threshold is given beforehand or computed from the given set of data. Moreover different thresholds may be applied to different classes of meshes in order to preserve the density variations across different regions of a road network. If the density of a mesh goes beyond the threshold, then there is a need to eliminate one or more road segments on its boundary.

4.2 Determination of thresholds for mesh density

There are two different ways to determine a density threshold for each class of meshes. They are respectively based on theoretical analysis, map specifications and empirical study.
For maps at a certain scale, there must be a minimum size for a mesh unit below which the mesh cannot be perceived anymore. Correspondingly, there must be a threshold of mesh density beyond which one or more segments on the boundary of the mesh must be eliminated so that two or more adjoining meshes could be merged to form a larger mesh. Such a threshold is regarded as the permissible largest density (PLD), which implies the longest possible length of roads in the smallest visible area. The minimum mesh size could be set to the SVO (Smallest Visible Object) in "natural principle" proposed by Li and Openshaw (1992, 1993). SVO could be a small circle, a raster cell or any other geometric entity. So the PLD i.e. density threshold can be expressed in terms of the ground size and the principles of the SVO as follows:

$$D_{th} = 4\big/\left(S_t \cdot L_m \cdot (1 - S_s / S_t)\right) \qquad (2)$$

Sometime other values similar to the mesh density are required for area symbols in cartographic specifications and the mesh density can be computed according to them. For example, the mesh area has been also mentioned with the threshold in cartographic specifications. Since the mesh area as a constraint does not take the shape of the mesh into account, it can only lead to the elimination of part of meshes. So the threshold of density can be computed by this area specified.

The second way is based on the statistic of the sample maps to determine the threshold of density. For example, an empirical study was carried out for the scale change from 1:10,000 to 1:50,000. Two sample maps are from the urban street network, which mostly contains SS (Secondary Streets). The diagram in Figure 5 describes the density distributions of SS meshes at different scales. Towards the side with a larger density (between 0.026 and 0.070), the map at 1:10,000 contains more meshes than the map at 1:50,000 for each density value, which indicates that the number of meshes needs to be reduced with the decreasing map scale. In other words, some road segments should be deleted and their associated meshes merged. However towards the side of smaller density (between 0.010～0.026), the map at smaller scale contains more meshes than that at larger scale for each density value. It indicates that the meshes with larger densities have been merged, thus leading to a reduced number of meshes with smaller densities. The density value 0.026 is a critical point. That is, most of the meshes with a density over 0.026 will not be represented on the map at 1:50,000 and need selective omission. This critical point might be used as the density threshold for SS meshes at scale 1:50,000.

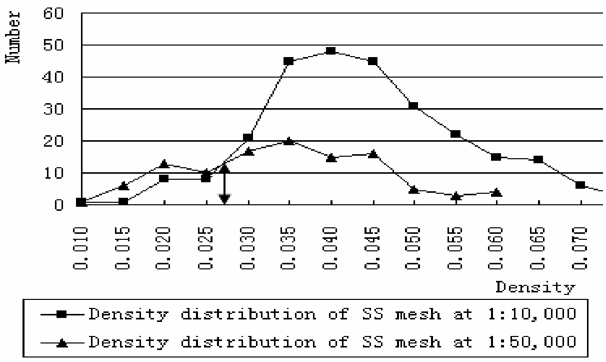


Figure 5. Threshold determination of mesh density by statistics of sample maps

The threshold values for different types of road meshes can be determined with different methods according to the ranking. The density thresholds for the meshes with lower classes will be set to 0, meaning the elimination of all these meshes at smaller scale. The density thresholds for the roads with high classes are set to the PLD in order to retain them all. The density thresholds for meshes with the middle classes can be determined using the second way.

4.3 Elimination of road segments and merging of adjoining meshes

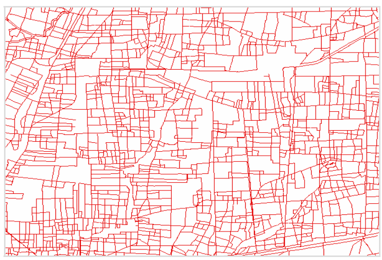
The aggregation of adjoining meshes must start from those candidates with the largest density. This requires that meshes with densities beyond the threshold value be identified as candidates and sorted by density. Doing so is to protect the meshes with densities beyond the threshold value from being over-merged, which may cause some road segments to be unnecessarily deleted.

In order to determine which segments to be eliminated, some parameters need to be adopted. The attributes such as name, class and length of road segment may form a set of such parameters. However, the values of these attributes are not always available. In this methods, two further parameters are considered, i.e. the length and the degree of stroke (Thomson, 2006) because they reflect well the geometric, semantic and topological characteristics of road segments. The parameters reflecting importance of road segments have a priority order for road selection. The following parameters are used, i.e. road class (C), degree of stroke (Ds), length of stroke (Ls), length of segment (L), with the importance in descending order. For example, if there is only one road segment of the lowest class on the mesh boundary, then it will be definitely eliminated. During the elimination process, a number of rules should be created based on the aforementioned parameters. Some of these rules used in this study are formulated as follows:

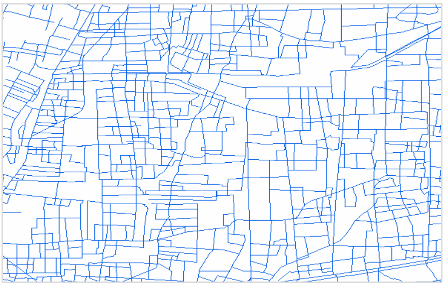
Rule 1: The road segments of the lowest class should be eliminated;
Rule 2: If there are some road segments with Ds equal to 1, then eliminate the shortest one of them;
Rule 3: If there is no road segments with Ds equal to 1, then search for road segments which have lengths below the threshold Ls and eliminate the shortest one of them;
Rule 4: A road segment connecting important features such as a dock will be preserved;
Rule 5: If the length of a stroke reaches a certain length, then the road segments on it will be preserved.

5. A CASE STUDY

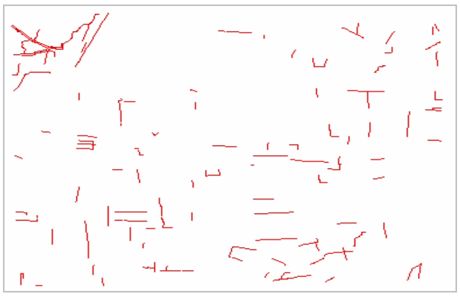
Based on the above methods of matching and selective omission, a software module was developed and integrated into the special toolkit for map updating by data generalization. It has been used to update 1:50,000 map roads by generalizing up-to-date 1:10,000 road data by the national bureau of surveying and mapping in China. A case is introduced in this section. The old road data at 1:50,000 and the new road data at 1:10,000 are shown in Figure 6a and 6b, and they cover the same area of about 400 km2. The updates retrieved are shown in Figure 6c, 6d and 6e respectively and the corresponding statistics are listed in Table 1. The results of data at 1: 50,000 updated are also represented in Figure 6. The results updated satisfy map specification and shows a higher efficiency and quality.



(a) the new road data at 1:10,000



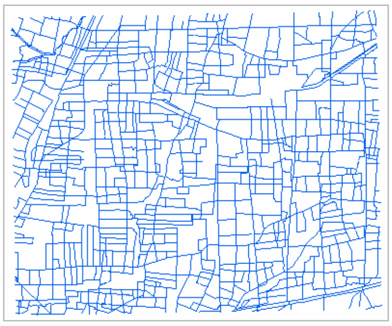
(b) the old road data at 1:50,000



(c) the new or changed roads

(d) the disappearance roads

(e) the generalization disappearance



(f) the road data updated at 1:50,000

Figure 6. Road data, updates and road data updated

Types of updates	Length	Percentage (item / updates)	Percentage (item/road length before updating)
New or changed	125.3	56.34%	11.97%
Disappearance	9.3	4.19%	0.89%
Generalization disappearance	87.8	39.47%	8.39%
Total	222.5	100%	21.25%
Total length of roads before updating	1047.2		
Total length of roads after updating	1075.4		

Table 1 Statistics for updates and road data (Length unit: Km)

REFERENCES

Badard, T. and Lemarié, C. Propagating updates between geographic databases with different scales, Chapter 10 of Innovations in GIS VII: *GeoComputation*, Atkinson, P. and Martin, D. (Eds.), Taylor and Francis, London, 1999.

Badard, T. Towards a generic updating tool for geographic databases . 1998, *In the proceedings of GIS/LIS'98 Annual Exposition and Conference*. Fort Worth, Texas, USA, 352-363.

Balley, S., Parent C and Spaccapietra, S. 2004, Modeling geographic data with multiple representations, *International Journal of Geographic Information Systems*, 18(4):329-354.

China Standardization Office, 2003, *Compilation of Standards for Mapping and Surveying-Map Making and Printing*, China Standardization Office, Beijing.

Devogele, T., Trevisan J. and Raynal L. 1996, Building a multi-scale database with scale transition relationships , In: Kraak, M.J., Molenaar M. and Fendel E. M., (eds), it Advances in GIS Research 2, Taylor & Francis, London. 337-351.

Filin, S. and Doytsher, Y. 1999,Linear approach to map conflation: matching of polylines, *Surveying and Land Information Systems*, 59(2):107-114.

Gabay, Y. and Doytsher, Y. Automatic adjustment of line maps , *Proceedings of the GIS/LIS'94 Annual Convention,* Arizona, Phoenix, USA, 1994, 1:333-341.

Hu Yungang, 2007, road data updating based on map generalization, doctoral dissertation, China.

Jones, C. B., Kider D.B., Luo L.Q., et al. 1996,Database design for a multi-scale spatial information system, *International Journal of Geographical Information Systems*, 10(8):901-920.

Jiang, B. and Harrie, L. 2004, Selection of streets from a network using self-organizing maps, *Transactions in GIS*, 8(3), 335-350

Kilpeläinen, T. and Sarjakoski, T. Incremental generalization for multiple representations of geographic objects, in: Lagrange J. and Weibel R., *GIS and Generalization*, Taylor & Francis, London, 1995,209-218.

Li, Z. L. and Openshaw,S. 1992, Algorithm for automated line generalization based on a natural principle of objective generalization, *International Journal of Geographical information Systems*, 6(5):373-389.

Li, Z. L. and Openshaw,S. 1993, A natural principle for objective generalization of digital map data, *Cartography and Geographic Information System*, 20:19-29.

Mackaness, W. A. 1995,Analysis of urban road networks to support cartographic generalization, *Cartography and Geographic Information Systems*, 22(4):306-316.

Muller, J. 1987,Optimum point density and compaction rates for the representation of graphic lines . *Proceedings AUTO CARTO8 (Baltimore)*. 221-230.

Rosen, B. and Saalfeld, A. 1985, Match criteria for automatic alignment, *Proceedings of Auto-Carto* VII, American Congress on Surveying and Mapping and American Society for Photogrammetry and Remote Sensing, 1-20.

Thomson, R. C. 2006, The 'stroke' concept in geographic network generalization and analysis. *Proceedings 12th International symposium on spatial data handling, Vienna*, 681-697.

Uitermark, H. Oosterom P. Mars, N. et al., 1998, Propagating updates: finding corresponding objects in a multi-source environment. In Poiker, T.and Chrisman, N. editors, *Proceedings of the 8th International Symposium on Spatial Data Handling*, Vancouver, Canada, 580-591.

Walter, W. and Fritsch, D. 1999, Matching spatial data sets: a statistical approach, *International Journal of Geographical Information Systems*, 13(5): 445-473.

Xiong, D. and Sperling,J. 2004, Semi-automated matching for network databases integration, *ISPRS Journal of photogrammetry and Remote sensing*, 59(1-2):35-46.